Validation of simulation models

Kleijnen, J.P.C.; Alink, G.A.

Publication date:
1992

KATHOLIEKE UNIVERSITEIT BRABANT
TILBURG UNIVERSITY

POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

VALIDATION OF SIMULATION MODELS: MINE-HUNTING CASE-STUDY

Jack P.C. Kleijnen
Gustav A. Alink

R11

FEW 537

Simulation Models

Refereed by Prof.dr.ir. C.A.T. Takkenberg

Kenmerk: 320.92.152

VALIDATION OF SIMULATION MODELS: MINE-HUNTING CASE-STUDY

JACK P.C. KLEIJNEN

Tilburg University, Tilburg, Netherlands

GUSTAV A. ALINK

TNO Physics and Electronics Laboratory, The Hague, Netherlands

February 1992

VALIDATION OF SIMULATION MODELS: MINE-HUNTING CASE-STUDY

JACK P. C. KLEIJNEN
Tilburg University, Tilburg, Netherlands

GUSTAV A. ALINK
TNO Physics and Electronics Laboratory, The Hague, Netherlands

Stringent validation requires that simulation and real-life responses have the same mean. The responses, however, may show not only sampling error but also measurement error. Moreover, simulated and real responses are not comparable if they are obtained under different environmental conditions or scenarios. Modules within the simulation model should be submitted to sensitivity analyses based on experimental design theory and regression analysis. A weaker validation procedure tests whether the estimated simulation and real responses are positively correlated (they do not necessarily have a common mean). These issues are illustrated through a study on mine hunting at sea by means of a sonar.

Validation in general is discussed in all textbooks on simulation; examples are Kleijnen and Van Groenendaal (1992), Law and Kelton (1991, pp. 298-324) and Pegden et al. (1990, pp. 133-162). These textbooks give many additional references; also see Reckhow (1989). In this paper, however, we study a specific case, namely a particular simulation model of mine hunting, called HUNTOP. This model has been developed in the Netherlands by the TNO Physics and Electronics Laboratory to be used by the Royal Netherlands Navy.

Other countries have similar simulation models for mine hunting.

In Section 1 we give the conceptual model of the search for mines on the sea bottom, which includes environmental factors (namely the mine field and the sea characteristics), the sonar system, the ship's course, and the human operator's performance. In Section 2 we validate the simulation model in two stages. In Subsection 2.1 we perform sensitivity analyses on some modules. The resulting regression 'metamodels' give encouraging results. In Subsection 2.2 we compare simulated detection probabilities resulting from the model as a whole, with real-life probabilities. We solve several statistical problems, such as dependencies between estimated detection probabilities of different mines. We emphasize the importance of measuring the environment or scenario that drives the simulation and the field test respectively. A complication is caused by measurement errors: around the assumed location of a mine a circle is draw and only detections within that circle are counted as detections. Several other measures of interest are briefly discussed. In Section 3 we discuss future research. For example, some specific input factors turn out to be important, and should be further investigated; until now some of them have been modeled rather poorly as qualitative factors. A complication is caused by the fact that 'false' detections may be counted as detections and 'true' detections may be ignored. A weaker validation procedure may test whether the estimated simulation and real detection probabilities are positively correlated (they do not necessarily have a common mean). In Section 4 we give conclusions; some apply to this particular case study only, whereas other conclusions hold for simulation and modeling in general.

1. MINE HUNTING: CONCEPTUAL MODEL

Mine hunting at sea is performed by ships that are equipped with sonar. Conceptually, a sonar beam may be viewed as a torchlight, that is, in the 'dark' a certain area becomes 'lighted' or 'insonified' so objects within that area may become visible on a sonar display. As the ship and the sonar move, new areas become visible, while previous areas move out of sight. Operationally, the total search area is divided into mutually exclusive and exhaustive strips. In the middle of each of these strips, an imaginary straight line is drawn, called the 'track'. The ship sails down one track, returns over the next track, and so on, until the whole area has been covered. Mines and NOMBOs (NOn-mine Minelike Bottom Objects) are dispersed over the area. These objects can only be detected if they are within the insonified area. The following discussion gives some information on the model that is presented in Figure 1.

INSERT FIGURE 1: Mine hunting model

Technically, sonar performance depends on sound velocity, which varies with water temperature and salinity. Obviously these characteristics vary with the water depth. It is standard to model the propagation of the sonar beam through the *Sound Velocity Profile* (SVP), which maps sound velocity as a function of depth. In the model the SVP is a simple piecewise-linear function, which is kept constant during the whole simulation run. In practice, the SVP varies from place to place, as the ship sails over the track. But even at the same place, the SVP will change due to seasonal and daily variations. We shall return to the SVP.

When an object is insonified, its echo appears on the sonar screen with a certain *contrast*. This contrast is determined by three components: (i) the echo of the object itself, (ii) the echo of the object's environment (reflections from sea bottom and surface), called '-reverberation', and (iii) acoustic noise (sounds are also generated by the ship, waves, and so on). The echo of the object depends deterministically on several factors, for example, on the 'aspect' angle at which the sonar beam hits the axis of a cylindrical object and on the object's dimensions. Reverberation depends deterministically on the 'grazing' angle at which the sonar beam hits the bottom and on the bottom type (for example, a rocky bottom reflects sound more than sand does). Acoustic noise may generate random, 'spurious' contrasts. Note that harmless objects may look like mines. Furthermore, mines may be hidden behind hills on the sea bottom: the bottom profile may be important.

A contrast may be missed by the *human operator*. Human behavior shows noise and is therefore represented by statistical distribution functions, called 'operator curves'. An operator curve gives the detection probability of an echo as an increasing function of the time that the echo has been visible. The model distinguishes three classes of 'object density': if there are many echoes then the detection probability of an individual object is lower. Besides object density classes the model distinguishes classes of sonar 'sector' angles; this angle determines the beam scanning pattern. Within each combination of object density and sector angle class there are four or five classes of contrast strength, each with its own operator curve. An object is visible only during a certain time, which depends on the sonar 'search window' (like the light circle of a torch) and on the position of the object relative to the ship's course. When the object becomes invisible or the operator is busy, the detection probability drops to zero. Whenever detection is made, the operator must classify the observed contrast as either a mine or a NOMBO. That classification may be true or false. The model, however, does not cover this classification stage nor any other follow-up operations such as sending an unmanned mini-submarine to identify a classified object or to neutralize or destroy the mine.

The laws of physics that govern sonar beam propagation are well known and are represented by deterministic relationships (for example, Snell's law). The environment, however, is not well known: accurate information on the current SVP and the sea bottom's profile is not available. The simulation uses a single SVP within one run. So, even if the model is perfect, it may give the wrong answer when it is fed with the wrong inputs. This type of uncertainty must be distinguished from random noise, which occurs in the operator module (and in other modules as we shall see); also see Kleijnen (1990).

Another problem is the continuous character of time in many physics laws. The simulation model, however, is programmed with time sliced into periods of fixed length (not of variable length as in discrete event simulation). In other words, the model consists of difference equations, not differential equations. The time slice has a length of three seconds if the ship's speed is two meters per second. Numerical accuracy is acceptable at this time step.

The model is 'calibrated', that is, there is one parameter that is used to modify the computed contrasts such that the model's outputs are closer to the outputs observed in practice; this parameter has no physical interpretation.

The model may be used for different *purposes*. In sensitivity analysis, different tactics for mine hunting are compared; for example, the 'tilt' angle of the sonar may be changed (in the torchlight analogy, we may think of shining farther away, so we see a larger area but with less intensity). Moreover, a given tactic may give different results depending on the environment; therefore the non-controllable factors should also be investigated. Besides the relative responses obtained in sensitivity analysis, absolute predictions are of interest: the expected detection probabilities in a given situation may be used to determine the 'huntability' of the mine field and to assess the performance of a particular sonar system.

The simulation model has nearly 40 inputs or factors, some of which were mentioned above. That model consists of a number of modules; for example, modules for the ship's position (which includes navigation error), the operator's state, the object's visibility, and the object's contrast; the latter three modules give the inputs for the detection probability module. Figure 1 shows the main modules of the model and their relations. For reasons of confidentiality we do not give more details on the model; those details are presented in Alink and Vermeulen (1991).

There are actually several options within the model. For example, the SVP may be either input to the model or it may be calculated as a function of salinity and temperature. So in Figure 1 SVP is not shown at the extreme left (input) but more to the right. We concentrate on the SVP as input. Other examples are reverberation and noise, which are also modeled in two ways. Moreover, there is an analytical variant of the mine hunting model.

Summarizing, in the model the mine detection probabilities depend primarily on the following factors.

(i) Environmental factors: the mine field (number of mine-like objects per square kilometer, mine orientation), the sea (depth, SVP, and noise level), and the sea bottom (type and profile).

(ii) The sonar system (technical specifications as well as operational settings such as tilt and sector angles).

(iii) The ship's course (including navigation error).

(iv) The operator's performance.

Note that the following report is more or less a chronological account of some issues that arose in the validation study.

2. VALIDATION

To validate the simulation model we proceed in two stages. In stage #1 we perform sensitivity analysis per module. The way we perform sensitivity analysis is better than the approaches followed in the three military case studies described in Fossett et al. (1991, p. 719). In stage #2 we compare simulated detection probabilities resulting from the model as a whole, with real-life probabilities.

2.1 Sensitivity Analysis per Module

Some modules give *intermediate* output that is impossible or hard to validate. We then apply sensitivity analysis per module. Because of time constraints we do not examine all but only two modules, namely the 'sonar window' and the 'visibility' modules.

The response variables of the *sonar window* module are the maximum and the minimum distances that relate to the area on the sea bottom insonified by the sonar beam. Those responses depend deterministically on several factors, namely SVP, water depth, and sonar parameters. SVP is a qualitative environmental factor, which we have already discussed. The sonar parameters are the tilt and sector angles (as we have already seen) plus the 'sideward' angle relative to the ship's course. The sonar rays hit the bottom under the grazing angle (which is determined by the SVP, the water depth, and the tilt angle).

As response variable we first take the minimum distance from the sonar to the insonified area on the sea bottom, denoted by y (actually the sonar position is projected onto the imaginary flat sea bottom). We investigate three inputs or factors: SVP or x_1, average water depth or x_2, and tilt angle or x_3. We specify a regression 'metamodel', which approximates the response as a simple function of the inputs (see Kleijnen, 1987). A second-order polynomial in x_2 and x_3 gives a multiple correlation coefficient R^2 between 0.96 and 0.98, for four different SVPs (we could also have applied cross-validation: see Kleijnen and Van Groenendaal, 1992). Qualitative knowledge about the simulated subsystem suggests that the regression coefficients have specific signs: $\beta_2 > 0$, $\beta_3 < 0$, and $\beta_{23} < 0$. The estimates turn out to have the correct signs; the pure quadratic effects are not significantly different from zero. For the second response, maximum distance, similar results hold, except for one SVP that results in an R^2 of only .68 and a non-significant β_2. We trust the underlying simulation module more if the estimated regression coefficients have the right signs, provided of course that the metamodel fits reasonably so the coefficients are not meaningless.

An object is *visible* if it is within the sonar window and not concealed by the bottom's profile. The bottom profile is modeled by a simple geometric pattern, namely hills of fixed heights with constant upward slopes and uniform downward slopes. A fixed bottom profile is used within a single simulation run. The orientation of these hills relative to the ship's course and to the direction of the sonar beam is also relevant: does the sonar look down a valley or is its view blocked by a hillside? The response variable of the visibility module is the time that the object is visible, expressed as a percentage of the time it would have been visible were the bottom flat (in which case no concealment could occur). This response is random because the ship's course is affected by navigation error. Navigation error is modeled by a normal distribution with the desired course over the track as the mean value. Because of this sampling error, several simulation runs are necessary to estimate the response.

We vary six inputs: water depth, tilt angle, hill height, upward hill slope, downward hill slope, and object's position on the hill slope (top, bottom, or in between). We keep the SVP and the orientation of the bottom profile constant and eliminate navigation error. We specify a quadratic metamodel for this module, and use a central composite design with 77 input combinations to estimate the 28 regression parameters. R^2 is .86 and the adjusted R^2 is .78. The upward hill slope has no significant effects (no main effect, no interactions with the other factors, no purely quadratic effect); these results agree with the qualitative knowledge of the system analysts.

2.2 Empirical versus Simulated Detection Probabilities

A 'run' is one voyage of the ship over the whole mine field. During that run a particular mine is either detected or not. So if M denotes the number of mines in the simulated mine field, and R the number of simulation runs, then

$$x_{ij} = 0 \quad \text{if simulated mine } i \text{ is not detected in simulation run } j \text{ with } i = 1,\ldots, M \text{ and } j = 1,\ldots, R;$$
$$= 1 \quad \text{if simulated mine } i \text{ is detected in simulation run } j. \tag{1}$$

This equation leads to the following definition of the detection probability p_i for mine i that holds for all runs:

$$P(x_{ij} = 1) = p_i;$$
$$P(x_{ij} = 0) = 1 - p_i. \tag{2}$$

Let K denote the number of field runs that are performed during a sea trial in a certain period. We suppose that the number of mines (namely M) in the simulation runs that are meant to validate the model, and in the field tests are equal. (After the validation we can change the number of mines in the model.) So

$y_{ik} = 0$ if real mine i is not detected in field run k with k $= 1, \ldots, K$;

$= 1$ if real mine i is detected in field run k. (3)

Analogous to (2) we define

$$P(y_{ik} = 1) = q_i;$$
$$P(y_{ik} = 0) = 1 - q_i. \quad (4)$$

A major problem in our case study is the inputs of the simulation model and the field test respectively. The SVP of the model is a crude approximation of the SVPs met in practice (in Section 1 we saw that the SVP in the model is a simple piecewise-linear function, which is kept constant during the whole simulation run). The real SVPs are poorly measured. The locations of the real mines in the field test are also not known exactly. So on one hand an echo is not counted as a detection if its origin is 'far' away from the assumed locations of the real mines. On the other hand 'false' echoes (NOMBOs and spurious contrasts) are counted as detections if their origins are close to (the assumed location of) a real mine. So environmental conditions like SVP are uncertain in the real world, and hence in the model. Obviously the detection probabilities in (2) and (4) depend on uncertain but deterministic inputs like the SVP and the mine locations and orientations. These inputs we call a 'scenario'. (These inputs must be distinguished from the stochastic inputs, namely navigation error, spurious contrasts, and human performance.) There are numerous scenarios (say) S_h with h = 1, 2, ... So analogous to (1) we define

$x_{ijh} = 0$ if simulated mine i is not detected in simulation
run j under scenario h with h = 1,2,...;
= 1 if simulated mine i is detected in simulation run j under scenario h. (5)

Analogous to (2) we define the conditional probabilities

$$P(x_{ij} = 1 | S_h) = p_{ih}. \tag{6}$$

To estimate p_{ih} from R simulation runs, we keep the scenario fixed at S_h and use pseudorandom numbers to sample navigation errors, spurious contrasts, and human operator performance. This estimator is denoted by $\hat{p}_{ih}$. Obviously

$$p_i = \sum_h P(\, x_{ij} = 1 \mid S_h \,)\, P(S_h) \neq P(\, x_{ij} = 1 \mid E(S)). \tag{7}$$

To estimate p_i (the average over all scenarios), we sample scenarios too. This estimator is denoted by $\hat{p}_i$. The scenario E(S) may be an impossible scenario, that is, a scenario that can never occur.

The validation procedure should not test the unconditional probabilities p_i and q_i, but the conditional probabilities p_{ih} and q_{ih}. Figure 2 illustrates how the detection probabilities - both the simulated and the real ones - may depend on the *scenario*; it includes confidence intervals for the estimators of the simulation and the field tests. The figure illustrates that both estimators may be sensitive to the scenario. At scenario S_1 the model is not valid; at S_2 it is valid, and at S_3 it may be considered acceptable. If, however,the model is run with scenario S_2

whereas the field test uses scenario S_1, then the model would be (incorrectly) rejected. So we must estimate how much the simulated and the real detection probabilities respond to the scenario. If the detection probabilities are found to be sensitive to the scenario, then scenarios must be measured accurately; otherwise only less stringent validation tests are possible. (The importance of the environment is also emphasized by Fossett et al. 1991, p. 714.) The figure illustrates that the estimated simulated probabilities $\hat{p}_{ih}$ may lie within the range of estimated field results $\hat{q}_{ih}$, so without measurement of the scenario the model cannot be rejected. Note that this figure may be compared with a queuing situation where the traffic load replaces the scenario and the average waiting time replaces the estimated detection probabilities: if the traffic load is not measured, it is virtually impossible to validate the model.

INSERT FIGURE 2: Sensitivity of detection probability to scenario, for mine i

In practice several field tests are run, each in a different period. Within each test several SVPs are measured. In our opinion it is wrong to compute the average detection probability based on the probability of a particular scenario. We found that some estimated simulated detection probabilities are not sensitive to the scenario: these probabilities are always zero for some mines and one for some other mines, whatever the scenario is.

Technically, we specify the hypothesis that, for each mine, the model and the real system give the same detection probability under scenario h:

$$H_0\colon\ p_{ih} = q_{ih}. \tag{8}$$

Both probabilities can be estimated, one by simulation and one by field runs, assuming that the scenario can be fixed at S_h. The resulting estimator $\hat{p}_{ih}$ is binomially distributed (with parameters p_{ih} and R), since the simulation runs give independent responses: x_{ijh} and $x_{ij'h}$ are independent for $j \neq j'$ with $j, j' = 1, \ldots, R$ (mine detections within the same run may be dependent, as we shall see after equation 15). So

$$var(\hat{p}_{ih}) = p_{ih}(1-p_{ih})/R. \tag{9}$$

The field runs give similar binomial variables $\hat{q}_{ih}$ with parameters q_{ih} and K. Because the simulation outputs depend on pseudorandom numbers whereas the real outputs depend on other random events, the estimators $\hat{p}_{ih}$ and $\hat{q}_{ih}$ are independent. To test the null-hypothesis (8), we derive the variance of $\hat{p}_{ih} - \hat{q}_{ih}$ under the hypothesis that the simulated and the real probabilities equal (say) r_{ih}:

$$\begin{aligned} var(\hat{p}_{ih} - \hat{q}_{ih} | \, p_{ih} = q_{ih} = r_{ih}) &= r_{ih}(1 - r_{ih})/R + r_{ih}(1 - r_{ih})/K \\ &= r_{ih}(1 - r_{ih})(K + R)/(R\,K). \end{aligned} \tag{10}$$

To estimate the common parameter r_{ih} we propose the weighted average

$$\hat{r}_{ih} = \hat{p}_{ih}\,R/(R + K) + \hat{q}_{ih}\,K/(R + K). \tag{11}$$

Using

$$E(\hat{r}_{ih}^2) = var(\hat{r}_{ih}) + [E(\hat{r}_{ih})]^2 = r_{ih}(1 - r_{ih})/(R + K) + r_{ih}^2, \tag{12}$$

we derive the unbiased estimator of the variance in (10):

$$\hat{var}(\hat{p}_{ih} - \hat{q}_{ih} | \, \hat{p}_{ih} = \hat{q}_{ih} = r_{ih}) = \hat{r}_{ih}(1 - \hat{r}_{ih})(K + R - 1)/(R\,K). \tag{13}$$

Obviously the hypothesis (8) requires a two-sided test. Because there are several mines (M > 1), we apply Bonferroni's inequality. So the (composite) null hypothesis is rejected if one or more mines give significantly different estimated detection probabilities, each mine being tested at an individual type I error rate of α/M where α denotes the 'experimentwise' type I error rate (see Kleijnen, 1987 or Miller, 1981). For convenience we use the Gaussian distribution to approximate the distribution of the difference between two binomial variables. Its estimated mean is

$$\hat{p}_{ih} - \hat{q}_{ih} = \sum_{j=1}^{R} x_{ijh}/R - \sum_{k=1}^{K} y_{ikh}/K. \tag{14}$$

Its estimated variance was given in (13). So if z_α denotes the 'upper α point' or $1 - \alpha$ quantile of the standard normal distribution, we reject the null-hypothesis (8) if

$$\max \left[| \hat{p}_{ih} - \hat{q}_{ih} | / \sqrt{\hat{var}(\hat{p}_{ih} - \hat{q}_{ih} | \; p_{ih} = q_{ih} = r_{ih})} \;\right] > z_{\alpha/M}. \tag{15}$$

We emphasize that Bonferroni's inequality also applies if the M estimated probabilities within a given simulation run are dependent. Indeed, if the operator is busy with one mine then there is a higher chance that he misses the next mine. Similarly the estimated probabilities for various mines within a particular field test may be dependent.

Note that Brennan and Quan (1990) give exact confidence intervals for a single binomial parameter (p or q), not their difference (p-q). They use not the Gaussian approximation, but the original binomial distribution. Moreover, they do not follow the traditional approach that accounts for the discrete character of the binomial distribution and give conservative confidence intervals. Instead they follow a Bayesian approach, assuming no prior information on the binomial parameter. Also see Louis (1981) for the special case of observing no successes (x or y equal zero). Application of these techniques to our case study deserve more research.

When testing the validity of a model, there are two classical error sources, namely the type I or α error and the type II or β error:

α =probability of rejecting the model if the model is valid;

β =probability of accepting the model if the model is not valid. (16)

The complement of the type II error probability, $1 - \beta$, is called the 'power' of the test. That power increases as the model specification error $\delta_{ih} = |p_{ih} - q_{ih}|$ increases. Bonferroni's inequality implies that the 'experimentwise' type I error rate is α (as we have already mentioned below equation 13). The type II error probability increases as the type I error probability decreases, given fixed sample sizes (R and K). A classical value for the experimentwise error rate α is 0.20 (this means that the per comparison error rates are α/M). So, given the sample sizes R and K, the type I error probability α, and the model error δ, we can compute the β error probability. To decrease both error probabilities we can increase the sample size of the simulation; the sample size of the field test is usually given.

A complication is caused by the *measurement errors* in the field tests. In these tests a circle with a given radius is drawn around the location of the mine, assuming that location is exactly known. For validation purposes, the mine is supposed to be detected if and only if the operator records a contrast within that circle. Consequently, if the operator sees a false contrast (minelike object or spurious contrast) that falls within the circle, that echo is counted as a detection. On the other hand, a detection may be recorded outside the circle, and then it is not counted.

The statistical analysis outlined above, deviates from the analyses that is used in certain (confidential) naval studies (one of the latter analyses is performed by Van Zeebroeck).

Above we concentrated on the detection probabilities of individual mines (there are M mines). Related *measures of interest* are the average detection probabilities per 'strip'; there are several parallel strips (at fixed athwart distances) on both sides of the tracks, which may contain several mines. The detection probabilities are usually assumed to be the same for all mines within the same strip. The probabilities per strip can be further aggregated into the overall detection probability for the whole mine field. Naval experts are interested in the 'characteristic detection width' and the 'characteristic detection probability', denoted by A and B. These quantities are derived from the function that expresses the detection probability p as a function of (say) v, the athwart distance of the mine to the track. The function p(v) generally decreases as v increases. The following equations yield A and B:

$$W = \int_{-\infty}^{\infty} p(v)\, dv; \quad \int_{-W_1}^{W_1} p(v)\, dv = (2/3)W; \; A = 3\, W_1; \; B = W/A. \qquad (17)$$

We shall return to these measures in Section 3.

We do not present the results of the validation tests, as these results are confidential and based on procedures that require further justification.

3. FUTURE RESEARCH

Sensitivity analysis was applied to only two modules. So not all 40 factors of the total model have been investigated systematically. Such *factor screening* can be done through the screening method presented in Bettonvil and Kleijnen (1991). Examples of factors deserving further research are the length of time the operator is busy processing a detected object or a spurious contrast, the order of processing objects in the simulation program, the 'reverberation' function (reverberation affects contrast; reverberation is a function of range and 'grazing' angle), the magnitude of the navigation error's variance, certain factors in the formulas used to compute acoustic noise, the way operator curves that hold for a fixed contrast are used, the calibration factor.

A factor that certainly requires more research is the *SVP*. In practice that factor is hard to measure sufficiently since it depends on time and place. In the model it is treated as a qualitative factor. Such a nominal scale indicates lack of knowledge. Moreover, the simulation uses a single SVP per run. One challenge is to find 'robust' operating procedures, that is, procedures that are not sensitive to the specific SVP. It would also be useful to develop a real-time measurement device installed on board of the ship. Its measurements provide time and space dependent input to the simulation model, which then becomes a decision support system (DSS). The Royal Netherlands Navy have acknowledged this need and have proceeded to acquire such a system.

Bottom profile is also a qualitative factor. In the model a simple geometric pattern is used, whereas in practice the bottom may be erratic (fractiles might be used to model that profile more realistically). Moreover bottom type (mud, sand, rock, etc.) is modeled crudely. Bottom type is scaled from one to four, whereas it is actually a qualitative factor.

Sensitivity analysis should be applied to find out which inputs are really important. Collecting information on those inputs deserves much effort. If nevertheless it is impossible or impractical to collect reliable information on those inputs, then *risk analysis* may be applied. A probability distribution of inputs is then derived from the users' expert knowledge, which yields a probability distribution of output values; see Kleijnen (1990). Whether to apply sensitivity or risk analyses requires more research.

The simulated normally distributed *navigational error* was found not to have the desired mean. This error may be modeled better by specifying positive correlation. So if y_t denotes the actual position at time t and e_t the navigational error at that time, then

$$y_t = \mu + e_t \ , \tag{18}$$

where the error forms a time series

$$e_t = \rho \ e_{t-1} + z_t, \tag{19}$$

where ρ is the (positive) autocorrelation coefficient and z_t denotes a normally independently distributed variable with zero mean and variance (say) σ_z^2 such that e_t has the prespecified variance σ_e^2 (see Kleijnen and Van Groenendaal, 1992). This model implies that the ship's position is a weighted average of the desired course μ and the previous position y_t, augmented with an independent normal error with zero mean:

$$y_t = (1 - \rho\ )\ \mu + \rho\ y_{t-1} + z_t. \tag{20}$$

We explained that a circle with a given radius is drawn around the 'true' location of the mine, causing *measurement errors* (if and only if the recorded location is within the circle, the mine is supposed to be detected). We might refine this procedure by giving higher weights to any recorded detection, the closer it lies to the true position of a mine. (The weight function could be some bivariate distribution with means equal to the true coordinates and with such a shape that the weights decrease as specified by the naval experts; for example, with 90% probability a mine is counted as being detected if a recorded object lies no more than 20 meters from a true location; multivariate distributions of many shapes are surveyed in Johnson, 1987). Until now weights were zero or one. Actually we propose the following approach.

The current version of the model ends at the stage of mine detection, that is, it does not include the follow-up operations of classification and destruction of mines. Consequently, any contrast that the operator interprets as a mine (even if that detection is caused by a minelike object or a spurious contrast) and that is 'close' to an actual mine, is a success. Of course, calling 'mine!' all the time would generate a success probability of one, but would also waste much time and energy in the follow-up phase (which is not modeled). Therefore we should separately measure true detections (caused by whatever echo close to the true location of a mine) and false detections (caused by minelike objects and spurious contrasts only). (These detections resemble the type I and II errors in hypothesis testing; see eq. 16.) True and false detections should be measured, not only in the field tests but also in the simulation. In the current model, however, spurious contrasts are never counted as successes.

To estimate the *characteristic values A and B*, the estimated detection probabilities are processed after the simulation has been finished. All M estimated probabilities of a particular field test $\hat{q}_i$ and their athwart distances v_i are collected, ignoring measurement errors of v. Since this approach further ignores the scenario h (for example, the SVP), the resulting cloud of M observations (v_i, $\hat{q}_i$) is very erratic. We propose to estimate $p_h(v)$ from the estimated probabilities *per* scenario; for example, a mine farther away from the ship has a smaller detection probability, *given* a certain SVP. This estimation is possible provided a real-time measurement device is installed on board of the ship. Alternatively we can aggregate over scenarios and estimate p_i (which was defined in equation 7). This yields M pairs (v_i, $\hat{p}(v_i)$). These observations can be used to estimate the function p(v). The estimation of A and B through (17) may use nonlinear regression analysis, which yields estimates and confidence intervals. Note that we do not aggregate detection probabilities over strips since that aggregation means loss of information; moreover the width of the strip is subject to discussion. To validate the simulated characteristic values A and B we should measure the actual scenarios in the field tests, as we mentioned in the discussion of Figure 2. Note that the estimators of A and B are negatively correlated (B = W/A; see eq. 17).

We may formulate a *less stringent validation* requirement than we did in the composite null-hypothesis in (8): the simulated and the real detection probabilities are not necessarily equal, but the estimated simulation and real probabilities are *positively correlated*. So if a mine has a relatively high estimated detection probability in reality, then the estimated simulation probability should also be relatively high. To test this hypothesis we formulate the regression model

$$\hat{p}_{ih} = \beta_{0h} + \beta_{1h}\,\hat{q}_{ih} + \epsilon_{ih}, \tag{21}$$

where ε_{ih} is assumed to be 'white noise' (normally independently distributed with mean zero and variance σ_h^2). So if scenarios are measured, then we can plot $\hat{p}_{ih}$ as a function of $\hat{q}_{ih}$, and use ordinary least squares to estimate the intercept and slope of the straight line that passes through the 'cloud' of points. The null-hypothesis is

$$H_0\colon \beta_{1h} \leq 0. \tag{22}$$

To test this hypothesis we use the standard t statistic. So we reject the null-hypothesis in (22) and accept the simulation model if there is strong evidence that the estimated simulation and real detection probabilities are positively correlated; see Kleijnen and Van Groenendaal (1992).

The number of points in the 'cloud' would be M if the scenario could be kept constant during the whole field test. Several field tests may be combined to get more observations, provided the scenarios are measured. If scenarios are not measured, then collecting all data in a single diagram creates extra noise. Technically it means that the index h is deleted in (21) and (22).

The weaker validation requirement makes sense if the model is used to predict relative responses (as in sensitivity analysis of tactics and sonar design), not absolute responses (needed to gauge the 'huntability' of a mine field). In the latter situation, input data of higher accuracy are necessary.

The model may be augmented with mine *classification and destruction* procedures. User requirements may be better satisfied if *animation* is applied to present the simulation model; see Kleijnen and Van Groenendaal (1992).

4. CONCLUSIONS

The conceptual model of the search for mines includes environmental factors (mine field and sea characteristics), the sonar system, the ship's course, and the human operator's performance. The conceptual validity of the model seems good.

We validated the simulation model in two stages. First we performed sensitivity analysis for some modules, namely the sonar 'window' module and the object visibility module. The resulting regression metamodels corroborate the modules' validity.

Next we compared simulated detection probabilities resulting from the model as a whole, with real-life probabilities. We derived a statistic to test that estimated simulated and real probabilities have the same mean. We emphasized the importance of measuring the environmental scenarios that drive the simulation and the field test respectively.

Finally we proposed several improvements. Screening all factors is technically feasible. Some specific factors have already been found to be important, but should be further investigated. Examples are sound velocity profile or SVP, bottom type, and bottom profile. Until now these factors have been modeled and measured rather poorly. It would be useful to develop a real-time measurement device, to be installed on board, which would feed time and space dependent SVPs to the simulation model, which then becomes a true DSS. If, however, objective information on important factors cannot be obtained, then risk analysis may be applied. Navigation error may be modeled better as a time series with negative autocorrelation. True detections (caused by whatever echo close to the true location of a mine) and false detections (caused by minelike objects and spurious contrasts only) should be measured separately. Other responses of interest to naval experts, namely the characteristic detection width and probability, should be better estimated. A weaker validation procedure may be applied to test that estimated simulation and real responses are positively correlated (they do not necessarily have a common mean). The latter test is valid if only relative responses are important; this is the case in sensitivity analyses (of the total simulation model) comparing different tactics and sonar parameters. Absolute probabilities are needed to determine the 'huntability' of a mine field. Improvements of the current model should make it possible to eliminate an artificial calibration parameter (introduced to get better fit between simulated and field test results). Animation may be used to get naval experts involved in the model construction, verification, validation, and operational implementation. The simulation results should be used to validate an analytical model that is also available for mine hunting.

REFERENCES

ALINK, G.A., and J.F.J VERMEULEN, 1991. *Validation of the Mine Hunting Model HUNTOP*. Report no. FEL-91-A096 (confidential, except for abstract), TNO Physics and Electronics Laboratory, The Hague.

BETTONVIL, B., and J.P.C. KLEIJNEN 1991. *Identifying the Important Factors in Simulation Models with Many Factors*. Tilburg University, Tilburg, Netherlands.

BRENNER, D.J., and H. QUAN, 1990. Exact confidence inter vals for binomial proportions - Pierson and Hartley revisited. *Statistician* **3**, 391-397.

FOSSETT, C.A., D. HARRISON, H. WEINTROB, and S.I. GASS, 1991. An assessment procedure for simulation models: a case study. *Opns. Res.* **39**, 710-723.

JOHNSON M.E. 1987. *Multivariate Statistical Simulation*. Wiley, New York.

LAW A.M., and W.D. KELTON, 1991. *Simulation Modeling and Analysis; Second Edition*. McGraw-Hill, New York.

KLEIJNEN, J.P.C. 1987. *Statistical Tools for Simulation Practitioners*. Dekker, New York.

KLEIJNEN, J.P.C. 1990. Statistics and deterministic simula tion: why not? *Proceedings of the 1990 Winter Simulation Conference,* edited by O. Balci, 344-346.

KLEIJNEN, J.P.C., and W. VAN GROENENDAAL, 1992. *Simulation: a Statistical Perspective*. Wiley, Chichester, United Kingdom.

LOUIS, T.A. 1981. Confidence intervals for a binomial para meter after observing no successes. *Am. Stat.*, **35**, 154.

MILLER, R.G. 1981. *Simultaneous Statistical Inference;*

Revised Edition. Springer-Verlag, New York.

PEGDEN C.P., and R.E. SHANNON and R.P. SADOWSKI. 1990. *Introduction to Simulation using SIMAN*. McGraw-Hill, New York.

RECKHOW, K.H. 1989. Validation of simulation models: philosophical and statistical methods of confirmation. In *Systems & Control Encyclopedia: Theory, Technology, Applications*., edited by M.G.Singh, Pergamon Press, Oxford, 5011-5015.

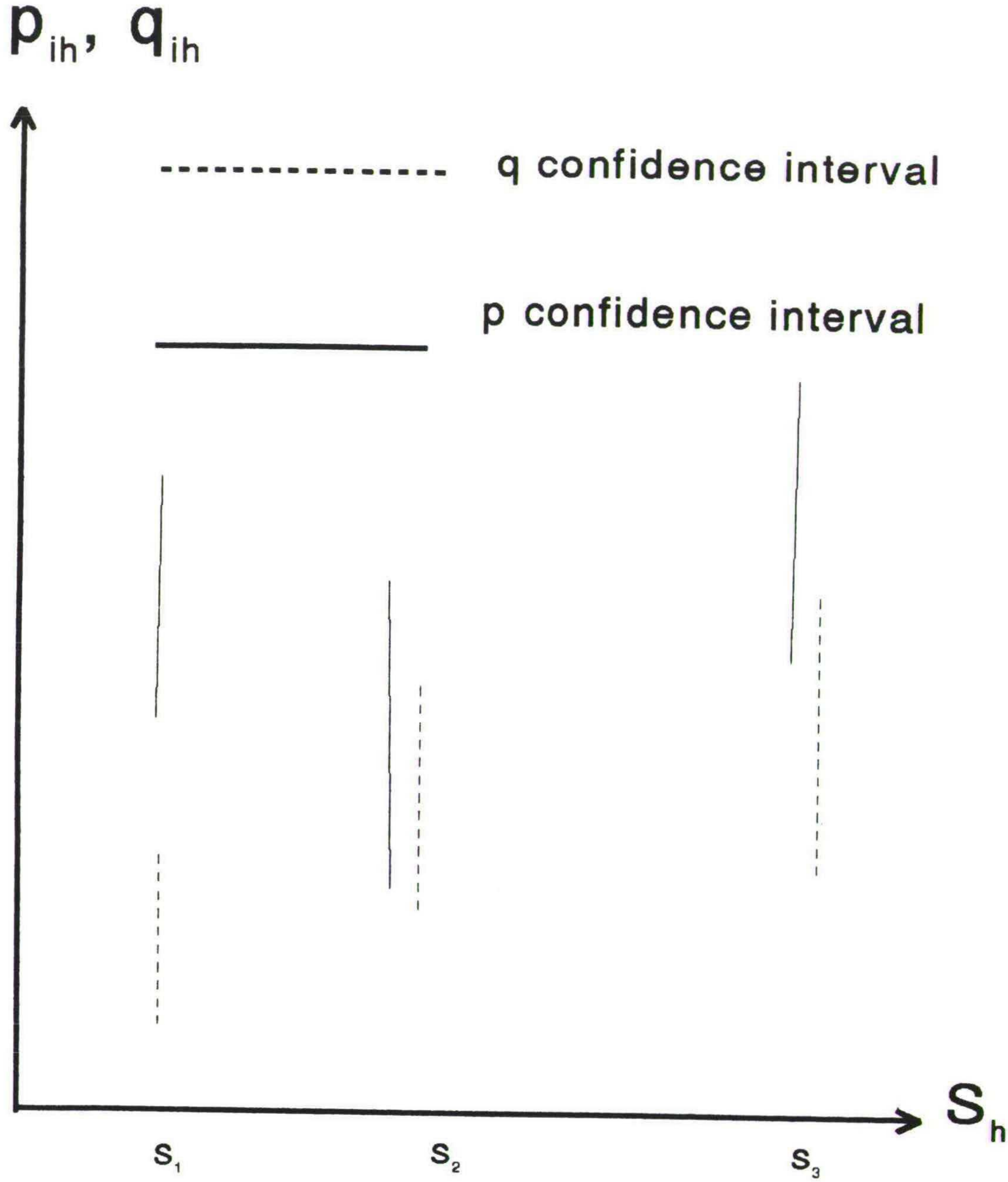
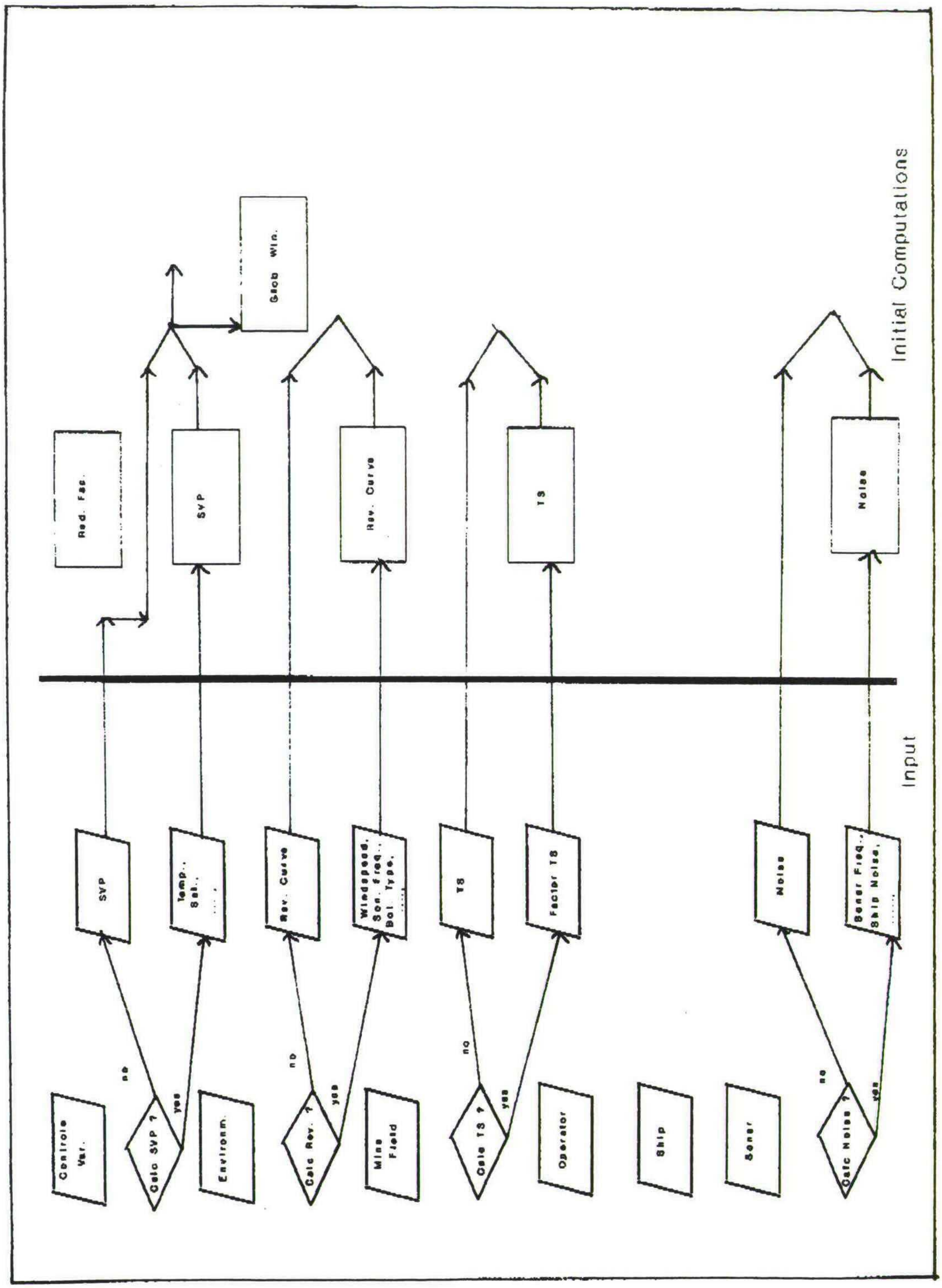
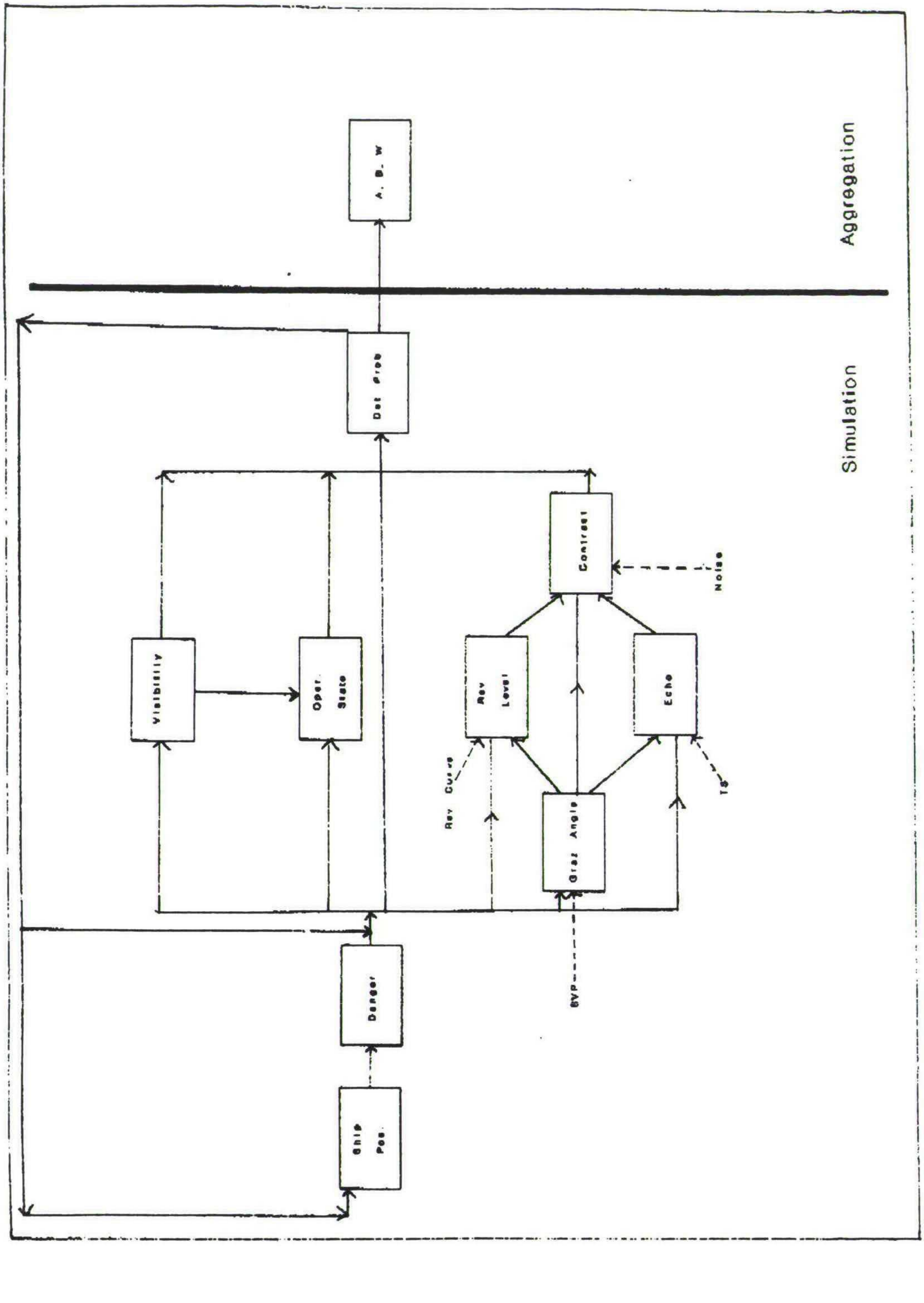


Figure 2. Sensitivity of detection probability to scenario, for mine i


Initial Computations
Input
Glob Win.
Red. Fac.
SVP
Rev. Curve
TS
Noise
SVP
Temp., Sal., ...
Rev. Curve
Windspeed, Son. freq., Bot. Type, ...
TS
Factor TS
Noise
Sonar Freq., Ship Noise, ...
Controls Var.
Calc SVP ?
Environm.
Calc Rev. ?
Mine Field
Calc TS ?
Operator
Ship
Sonar
Calc Noise ?
no
yes



Aggregation
Simulation
A. B. W
Det Prob
Visibility
Oper. State
Contrast
Noise
Rev Level
Echo
Rev Curve
TS
Graz Angle
SVP
Danger
Ship Pos.

IN 1991 REEDS VERSCHENEN

466 Prof.Dr. Th.C.M.J. van de Klundert - Prof.Dr. A.B.T.M. van Schaik
Economische groei in Nederland in een internationaal perspectief

467 Dr. Sylvester C.W. Eijffinger
The convergence of monetary policy - Germany and France as an example

468 E. Nijssen
Strategisch gedrag, planning en prestatie. Een inductieve studie binnen de computerbranche

469 Anne van den Nouweland, Peter Borm, Guillermo Owen and Stef Tijs
Cost allocation and communication

470 Drs. J. Grazell en Drs. C.H. Veld
Motieven voor de uitgifte van converteerbare obligatieleningen en warrant-obligatieleningen: een agency-theoretische benadering

471 P.C. van Batenburg, J. Kriens, W.M. Lammerts van Bueren and R.H. Veenstra
Audit Assurance Model and Bayesian Discovery Sampling

472 Marcel Kerkhofs
Identification and Estimation of Household Production Models

473 Robert P. Gilles, Guillermo Owen, René van den Brink
Games with Permission Structures: The Conjunctive Approach

474 Jack P.C. Kleijnen
Sensitivity Analysis of Simulation Experiments: Tutorial on Regression Analysis and Statistical Design

475 C.P.M. van Hoesel
An $O(nlogn)$ algorithm for the two-machine flow shop problem with controllable machine speeds

476 Stephan G. Vanneste
A Markov Model for Opportunity Maintenance

477 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts
Coordinated replenishment systems with discount opportunities

478 A. van den Nouweland, J. Potters, S. Tijs and J. Zarzuelo
Cores and related solution concepts for multi-choice games

479 Drs. C.H. Veld
Warrant pricing: a review of theoretical and empirical research

480 E. Nijssen
De Miles and Snow-typologie: Een exploratieve studie in de meubelbranche

481 Harry G. Barkema
Are managers indeed motivated by their bonuses?

482 Jacob C. Engwerda, André C.M. Ran, Arie L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^T X^{-1} A = I$

483 Peter M. Kort
A dynamic model of the firm with uncertain earnings and adjustment costs

484 Raymond H.J.M. Gradus, Peter M. Kort
Optimal taxation on profit and pollution within a macroeconomic framework

485 René van den Brink, Robert P. Gilles
Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures

486 A.E. Brouwer & W.H. Haemers
The Gewirtz graph - an exercise in the theory of graph spectra

487 Pim Adang, Bertrand Melenberg
Intratemporal uncertainty in the multi-good life cycle consumption model: motivation and application

488 J.H.J. Roemen
The long term elasticity of the milk supply with respect to the milk price in the Netherlands in the period 1969-1984

489 Herbert Hamers
The Shapley-Entrance Game

490 Rezaul Kabir and Theo Vermaelen
Insider trading restrictions and the stock market

491 Piet A. Verheyen
The economic explanation of the jump of the co-state variable

492 Drs. F.L.J.W. Manders en Dr. J.A.C. de Haan
De organisatorische aspecten bij systeemontwikkeling
een beschouwing op besturing en verandering

493 Paul C. van Batenburg and J. Kriens
Applications of statistical methods and techniques to auditing and accounting

494 Ruud T. Frambach
The diffusion of innovations: the influence of supply-side factors

495 J.H.J. Roemen
A decision rule for the (des)investments in the dairy cow stock

496 Hans Kremers and Dolf Talman
An SLSPP-algorithm to compute an equilibrium in an economy with linear production technologies

497 L.W.G. Strijbosch and R.M.J. Heuts
Investigating several alternatives for estimating the compound lead time demand in an (s,Q) inventory model

498 Bert Bettonvil and Jack P.C. Kleijnen
Identifying the important factors in simulation models with many factors

499 Drs. H.C.A. Roest, Drs. F.L. Tijssen
Beheersing van het kwaliteitsperceptieproces bij diensten door middel van keurmerken

500 B.B. van der Genugten
Density of the F-statistic in the linear model with arbitrarily normal distributed errors

501 Harry Barkema and Sytse Douma
The direction, mode and location of corporate expansions

502 Gert Nieuwenhuis
Bridging the gap between a stationary point process and its Palm distribution

503 Chris Veld
Motives for the use of equity-warrants by Dutch companies

504 Pieter K. Jagersma
Een etiologie van horizontale internationale ondernemingsexpansie

505 B. Kaper
On M-functions and their application to input-output models

506 A.B.T.M. van Schaik
Produktiviteit en Arbeidsparticipatie

507 Peter Borm, Anne van den Nouweland and Stef Tijs
Cooperation and communication restrictions: a survey

508 Willy Spanjers, Robert P. Gilles, Pieter H.M. Ruys
Hierarchical trade and downstream information

509 Martijn P. Tummers
The Effect of Systematic Misperception of Income on the Subjective Poverty Line

510 A.G. de Kok
Basics of Inventory Management: Part 1
Renewal theoretic background

511 J.P.C. Blanc, F.A. van der Duyn Schouten, B. Pourbabai
Optimizing flow rates in a queueing network with side constraints

512 R. Peeters
On Coloring j-Unit Sphere Graphs

513 Drs. J. Dagevos, Drs. L. Oerlemans, Dr. F. Boekema
Regional economic policy, economic technological innovation and networks

514 Erwin van der Krabben
Het functioneren van stedelijke onroerend-goed-markten in Nederland - een theoretisch kader

515 Drs. E. Schaling
European central bank independence and inflation persistence

516 Peter M. Kort
Optimal abatement policies within a stochastic dynamic model of the firm

517 Pim Adang
Expenditure versus consumption in the multi-good life cycle consumption model

518 Pim Adang
Large, infrequent consumption in the multi-good life cycle consumption model

519 Raymond Gradus, Sjak Smulders
Pollution and Endogenous Growth

520 Raymond Gradus en Hugo Keuzenkamp
Arbeidsongeschiktheid, subjectief ziektegevoel en collectief belang

521 A.G. de Kok
Basics of inventory management: Part 2
The (R,S)-model

522 A.G. de Kok
Basics of inventory management: Part 3
The (b,Q)-model

523 A.G. de Kok
Basics of inventory management: Part 4
The (s,S)-model

524 A.G. de Kok
Basics of inventory management: Part 5
The (R,b,Q)-model

525 A.G. de Kok
Basics of inventory management: Part 6
The (R,s,S)-model

526 Rob de Groof and Martin van Tuijl
Financial integration and fiscal policy in interdependent two-sector economies with real and nominal wage rigidity

527 A.G.M. van Eijs, M.J.G. van Eijs, R.M.J. Heuts
Gecoördineerde bestelsystemen
een management-georiënteerde benadering

528 M.J.G. van Eijs
Multi-item inventory systems with joint ordering and transportation decisions

529 Stephan G. Vanneste
Maintenance optimization of a production system with buffercapacity

530 Michel R.R. van Bremen, Jeroen C.G. Zijlstra
Het stochastische variantie optiewaarderingsmodel

531 Willy Spanjers
Arbitrage and Walrasian Equilibrium in Economies with Limited Information

IN 1992 REEDS VERSCHENEN

532 F.G. van den Heuvel en M.R.M. Turlings
Privatisering van arbeidsongeschiktheidsregelingen
Refereed by Prof.Dr. H. Verbon

533 J.C. Engwerda, L.G. van Willigenburg
LQ-control of sampled continuous-time systems
Refereed by Prof.dr. J.M. Schumacher

534 J.C. Engwerda, A.C.M. Ran & A.L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^{*}X^{-1}A = Q$.
Refereed by Prof.dr. J.M. Schumacher

535 Jacob C. Engwerda
The indefinite LQ-problem: the finite planning horizon case
Refereed by Prof.dr. J.M. Schumacher

536 Gert-Jan Otten, Peter Borm, Ton Storcken, Stef Tijs
Effectivity functions and associated claim game correspondences
Refereed by Prof.dr. P.H.M. Ruys